Introduction

Why Political Self-Deception?

The aim of this work is to contribute to the study and, possibly, the treatment of political deception.[1] More precisely, it intends to focus on a specific mode of political deception that is usually disregarded in political analysis, namely, self-deception (SD). Although it is a contested concept within philosophy and psychology, let us say, as a preliminary definition, that SD is the distortion of reality against the available evidence and according to one's wishes. The motivated distortion of data produced by SD obviously has significant consequences on the decision-making processes of political leaders, politicians, and government officials. Political decisions and policies induced by SD then lead to the deception of the public.

Political deception is generally acknowledged as a relevant, if problematic, issue for democratic politics.[2] The deception of the public is constantly denounced by media and by the press. It is mostly considered as intentionally achieved by governments and politicians, either by active lying – that is, by using false statements to mislead – or by intentional omission – that is, by withholding relevant information in order to

[1] When I talk of political deception, I refer to the public being deceived about something politically relevant. Deception is a success concept in that it means people have been made or come to believe something that it is false. This outcome is compatible with different modes of inducing the deception: (a) by intentional lying or intentional misleading; (b) by misperceptions and errors, (c) by SD, or (d) by having been unintentionally misled as the consequence of someone else's mistakes or SD. See, for example, J. E. Mahon, "The Definition of Lying and Deception," in *The Stanford Encyclopedia of Philosophy* (Spring 2016), Edward N. Zalta (ed.), http://plato.stanford.edu/archives/spr2016/entries/lying-definition/.

[2] The problem of political deception is widely dealt with in the literature. See, for example, N. Chomsky, *Necessary Illusions*: *Thought Control in Democratic Society*, Pluto Press, London, 1989; P. Rynard and D. P. Shugarman, eds., *Cruelty and Deception: The Controversy over Dirty Hands in Politics*, Broadview Press, Peterborough, ON, 1999; L. Cliffe, M. Ramsay, and D. Bartlett, eds., *The Politics of Lying*, Macmillan, London, 2000.

mislead, or by secrecy, propaganda, spin, and "bullshit," where the non-truth-oriented character of statements is quite explicit.[3] Both explanatory accounts and normative assessments of this widespread presence of deception have been proposed. They range from the realist position, holding that deception, secrecy, and manipulation are intrinsic to politics,[4] to the "dirty hands" position, justifying certain political lies under well-defined circumstances,[5] to the deontological stand denouncing political deception as a serious pathology of democratic systems.[6] Some recent works have more specifically focused on unpacking political deception, drawing distinctions among different kinds of lying – from above and from below; addressing either international or national audiences; for self-serving and for strategic reasons.[7] These works aim at understanding political deception without viewing it through the lens of moral outrage, yet, despite their more analytical and dispassionate approach, none entertains the possibility that political deception might partly be induced unintentionally by SD. Alternatively, it is sometimes conceded that the deception of the public is the by-product of government officials' (honest) mistakes.

In other words, political theory so far has considered political deception as induced either by lies, manipulation, and willful misinformation, or as the unintended consequence of illusions and misperceptions. The former calls for moral outrage and public exposure, the latter for cognitive analysis. But what if the false belief was candidly believed and, at the same time, the epistemic process of belief-formation was in the grip of an emotionally loaded desire switching on cognitive biases? Neither moral outrage nor purely cognitive analysis is of much help in this case, although political theory has basically oscillated from the first to the second.

This work aims at sidestepping the fork between pure dishonesty and cynicism, on the one hand, and honest mistakes, on the other. It aims to show that, more often than not, the misperception of reality, under various sources of psychological and emotional pressure, is driven by the desire to

[3] See H. Frankfurt, *On Bullshit,* Princeton University Press, Princeton, 2005.
[4] For example, M. Edelman, *The Politics of Misinformation*, Cambridge University Press, Cambridge, 2001.
[5] For example, M. Walzer "Political Action: The Problem of Dirty Hands," *Philosophy and Public Affairs*, 2, 1973, pp. 160–180.
[6] See M. Ramsay "Justification for Lying in Politics" (pp. 3–26); "Democratic Dirty Hands" (pp. 27–42); "Explanations: The Political Context" (pp. 43–55) in L. Cliffe, M. Ramsay, and D. Bartlett, eds. *The Politics of Lying.*
[7] M. Jay, *The Virtue of Mendacity: On Lying in Politics,* University of Virginia Press, Richmond, 2010; David Runciman, *Political Hypocrisy: Three Masks of Power from Hobbes to Orwell,* Princeton University Press, Princeton, 2010; John Mearsheimer, *Why Leaders Lie: The Truth about Lying in International Politics*, Oxford University Press, Oxford, 2011.

believe what one wishes to be the case, even if a dispassionate review of available data would lead any impartial observer to the opposite conclusion. Such a phenomenon is quite common, and evidence from experimental psychology seems to confirm that the common experience of seeing someone believing something against the evidence, but in accordance with his desires, has a scientific foundation.[8] Thus, there is no reason to believe that politics may be spared; moreover, SD is actually the best way to deceive others, for the more sincerely convinced a politician is, the more persuasive she appears to the public. Lastly, attributing the whole of political deception either to lying or to mistakes runs the risk of missing something important for explanatory as well as normative purposes.

I shall argue that SD represents a distinctive component in the wide realm of political deception, a component much unexplored, and yet worth examining in depth for its explanatory and normative bearings. Many students of politics have hinted that SD rather than straightforward deception might have been the case in the conduct and decision of political leaders; Hannah Arendt figures prominently among them in her renowned comment on *The Pentagon Papers*.[9] Few, however, have thoroughly pursued the hypothesis of political SD and analyzed it properly.[10] Moreover, the casual references to SD rely on the commonsense idea of SD, although its meaning is highly controversial and much debated in conceptual analysis. Therefore, SD often provides the umbrella for lumping together a variety of unjustified beliefs, such as myths and ideology. Obviously, SD cannot play any significant role in politics if its nature and meaning are not conceptually clear and distinct from other kinds of unwarranted beliefs and convictions.

There are both explanatory and normative reasons for considering the role of political SD seriously. The proper analysis of SD not only adds a missing piece to our knowledge of political deception but also provides a vantage point from which to explain political occurrences where public deception intertwines with political failure. It is often the case that the deception of the public goes hand in hand with faulty decisions. The

[8] See, for example, J. Lockards and D. Pahulus, eds., *SD: An Adaptive Mechanism*, Prentice Hall, Princeton 1988; V. S. Ramachandran, "The Evolutionary Biology of SD, Laughter, Dreaming and Depression: Some Clues from Anosognosia" *Medical Hypotheses*, 47 (1996): 347–362.

[9] Hanna Arendt, *Lying in Politics: Reflections on the Pentagon Papers, in Crises of the Republic*, Harcourt Brace Janovich, San Diego, 1972.

[10] Recent exceptions are Stephen Holmes in *The Matador's Cape. America's Reckless Response to Terror,* Cambridge University Press, Cambridge, 2007, and Robert Jervis, "Understanding Beliefs and Threat Inflation," in A. Trevor-Thrall and J. K. Cramer, *American Foreign Policy and the Politics of Fear: Threat Inflation since 9/11*, Routledge, London, 2009, pp. 16–39.

explanation of this fact divides current political studies into two groups. A first group discounts the deception as epiphenomenal of a slippery slope of mistakes and unintended consequences, thus exonerating politicians from moral responsibility. This view finds support in the strand of political psychology according to which misperceptions are simply the outcome of "cold biases." When applied to actual political failures, this view turns out to be too one-sided, and burdened with a clear exculpatory undertone, for the motivated and self-serving nature of such mistakes is left unaccounted for, and the associated responsibility is extenuated in epistemic inaccuracy. A second group focuses solely on the deceptive intent of politicians. But then, this conspiratorial position must also explain why such scheming liars plan so poorly, ending up exposed. Actually, there is no apparent causal connection between the government officials' deception of the public and the subsequent failure of the policy related to the public lie, whereas SD enables the analyst to account for (a) why the decision was bad, given that it was grounded on self-deceptive, hence false beliefs; (b) why the beliefs were not just false but self-serving, as in the result of the motivated processing of data; and (c) why the people were deceived, as the by-product of the leaders' SD. In addition to providing an explanation of the conjunction of bad planning, driven by self-serving motives, and public deception, the focus on SD will also imply an accurate analysis of the circumstances that set the process in motion.

This is the junction where normative reasons for considering SD become apparent. Mistakes and lies are detected by hindsight, and lies are usually detected only if unsuccessful. By contrast, a proper understanding of how SD works in crucial decision making opens up the possibility of identifying favorable circumstances for its taking place *ex ante;* hence, the possibility of devising preventive measures. SD is based on the cognitive distortion of data under the influence of some wish and not on bad will; yet SD is not the usual shortcut to believing what one wishes, unconstrained by evidence and epistemic processing. Only under special circumstances does motivation take the lead and bias the process of belief formation so as to produce the deceptive belief against the available evidence. Whereas bad will and mistakes can be neither predicted nor prevented, the circumstances conducive to SD can be detected. In principle, then, preventive strategies can be worked out at both the individual and the institutional level. As we shall see, SD is not under the control of the agent, but if it is acknowledged as an actual risk in policy-making, and if favorable circumstances are understood, we can at least count on a good starting point for working on prophylactic

measures: the assumption that no one likes to have duped him- or herself. In general, agents are not aware of entering a SD process, and external observers are the ones who usually detect others' SD. Putting together: (a) the typical circumstances in which SD may take place; and (b) the ability of external observers to identify other people's SD, a strategy of precommitment can be devised. Precommitment is a precautionary strategy, aimed at creating constraints to prevent people from falling prey to SD. If, *ex hypothesi,* independent observers have the authority to act as overseers of certain governmental decisions when the typical circumstances for SD are present, decision makers can become aware of the risk of certain chains of reasoning being biased, and might be offered the opportunity to reconsider their decisions. I am well aware of the many pragmatic difficulties in translating this idea into a viable institutional option. But the difficulties are no reason to dismiss the idea of preventive measures altogether.

A Clear Notion of SD

In order to appreciate the political role of SD, a preliminary analytical and critical understanding of its nature and working is required. As I noted earlier, the occasional hints at SD in political studies are too vague to be really useful, and the phenomenon is not really distinguished from a vast array of unwarranted beliefs and convictions that play an important part in political decisions and policies. Some of them are actually mistakes produced by cold biases, that is, by systematic cognitive distortions of our reasoning well described by cognitive psychologists. Others are motivated instead, induced by the effect on cognition of motivational states such as wishful thinking, illusions, and positive thinking. Then there are ideological convictions and political myths, projecting their fixed lens on data processing, that may or may not either cause or turn into self-deceptive beliefs. Political illusions, ideologies, and political myths, although not necessarily negative in decision making, are all contiguous with SD insofar as they imply a prejudiced consideration of data and induce unwarranted or not fully warranted beliefs. Only self-deceptive beliefs are, however, false by definition, being counterevidential, prompted by an emotional reaction to data that contradicts one's desires. If this is the specific nature of SD, as I shall show in the first part of this work, then self-deceptive beliefs are distinctly dangerous, for no false belief can ground a wise decision. It is therefore crucial to single out SD among the vast array of unjustified beliefs, motivated and unmotivated alike.

SD has been an issue in philosophy, especially in the Anglo-American tradition for the past forty years, and has more lately become a subject for experimental psychology and neuroscience. In the wide discussion on SD, there is no agreement on a standard notion or on its explanation, so much so that some scholars doubt that anything like SD may genuinely be the case. Skepticism about the phenomenon is voiced both by philosophers and by cognitive psychologists, albeit for different reasons. Philosophical skepticism is linked to the traditional view of SD as lying to oneself, which paves the way to paradoxes, for how can one be both the perpetrator and the victim of a lie? As a reaction to this paradoxical view, some scholars maintain that what has been described as SD is actually only a pretense to keep up one's image in front of other people.[11] But given that there is no need to pattern SD on the other-deception model, or to end up in paradoxes, as the current discussion has shown, then the philosophical skeptic has lost her footing for doubting SD as a genuine phenomenon. On the psychological front, skepticism for SD is voiced on the grounds that the more economical explanation of biased beliefs is available. According to this perspective, the presumption of any motivational influence on cognition is unnecessary, given that cold biases can do the job more directly. This reasoning, however, clashes both with many studies in experimental psychology showing that motivation often affects our perception, and with information gathering and processing in many different experimental settings. Given the rich phenomenology of SD and the growing experimental data on motivational interference on cognition, the reasons for doubting SD as a distinct phenomenon are very thin indeed.

Even conceding that SD is a genuine phenomenon, however, a plausible and persuasive account of SD is necessary before it is exported into politics. In this respect, philosophical discussion can serve three purposes in the economy of the present work. The first concerns conceptual clarity. This work's general aim is to single out the distinctive role of SD in the realm of political deception. If I employ an ambiguous notion, mainly relying on common sense, my argument is weakened thereby and exposed to all kinds of objections prompted by a more analytical and critical understanding of the phenomenon. There is no ready-to-use notion of SD; thus, the discussion and clarification of a viable conception

[11] M. Haight, *A Study on SD*, Harvester Press, Brighton, 1980; R. Audi, "SD, Action and the Will," *Erkenntnis*, 18, 1982: 133–158, E. Funkhouser, "Do the Self-Deceived Get What They Want?," *Pacific Philosophical Quarterly*, 86, 2005: 95–312.

of SD is not only intrinsically important but also constitutes a preliminary component of my project on political SD.

The second purpose that a critical analysis of SD is meant to serve relates to the issue of whether agents can be held responsible for their SD. I take it that a common reason for downplaying the role of SD in social and political life is a concern about responsibility. It may seem that introducing SD into the realm of political deception relieves politicians, government officials, and leaders deceiving the public wholly or partially of responsibility. If the deception of the public is explained by SD, instead of by a lie or by an intentional misleading by omission, then it is unclear whether moral blame and political outrage apply properly and fully in such a case, for the self-deceiver was neither aware of deceiving others as a consequence of his deception nor intending to deceive anyone else. The attribution of moral responsibility to self-deceivers is controversial: at most, it would seem that the responsibility would be reduced to negligence. But, if there were no possibility of holding people responsible for their SD, then SD would end up conflated with mistakes in the political domain. The fact that some mistakes are motivated does not significantly change the consideration that political analysts bestow on them. For my argument to stand, the issue of responsibility is thus crucial. This question, however, cannot be settled independently of the account of SD adopted: whether SD is seen as an intentional doing of the agent, albeit without full awareness and consciousness, or as a false belief causally induced by a motivational state that triggers cognitive biases has an obvious different bearing on the issue of responsibility.

Finally, the third purpose of philosophical and psychological analysis is to single out circumstances conducive to SD. As mentioned earlier, from a normative viewpoint, the advantage of SD lies in the hypothetical possibility of its prevention, and the latter, in turn, depends on the possibility of detecting when SD is likely to take place, which implies understanding its typical circumstances. The philosophy of SD has mainly disregarded the issue of circumstances, focusing instead on the definition of a puzzling concept and of its necessary and sufficient conditions. In other words, philosophical analysis has been concerned with defining criteria for identifying SD as a case of motivated irrationality. An account of SD that dispenses with its circumstances, however, is incomplete, for it cannot properly explain how SD strikes only selectively. Thus, my aim is to expand the scope of philosophical analysis so as to reach the circumstances, hence preparing the ground for prophylactic measures.

The first chapter of this book will be concerned with a critical analysis of the philosophical debate on SD. Among the many controversial issues in this debate, I shall consider one concerning whether SD is viewed as an intentional – although possibly unconscious – doing of the subject, or as a causal happening to the subject.[12] The first intentional view dominated the discussion until the nineties, whereas the second causal one is prevalent now. This issue has significant implications for the possibility of linking responsibility to SD; hence, settling the question of intentionality is crucial for the possibility of applying SD to politics. I shall therefore provide a critical presentation of the discussion on SD, both in its historical development and at its present stage, from the standpoint of the intentionality/causality debate. Intentionalists have so far failed to show how and why any agent would bring herself to believe what is knowingly false but corresponds to a wish of hers in a nonparadoxical way.[13] On the other side, the supporters of the causal/motivational view claim to have provided a simple, nonparadoxical, unified account of SD by showing that SD is the effect of a motivational state that causally triggers cognitive biases directly producing the false belief.[14] Although I share the causalists' criticisms, I nevertheless think that the intentional account embodies important intuitions that appear to be lost in the rival view. Briefly, the intentional account has had no difficulty in rendering SD as a *specific* form of motivated irrationality, well set apart from wishful thinking, positive thinking, or faith. Under the intentional description, the self-deceiver is not simply someone who believes what she wishes, but someone who has brought herself to believe that P that she (unconsciously) knows is false, for the true Non-P goes against her wish.[15] This paradoxical description captures a specific aspect of SD in the range of motivated irrationality, namely, the role played by the appraisal of the contrary evidence in allowing SD to start. This aspect is either overlooked or much diluted in the causal account, and the result is a loss of the specificity of SD, which is

[12] Anna Elisabetta Galeotti, "SD: Intentional Plan or Mental Event?" *Humana Mente,* 20, 2012, pp. 41–66.

[13] In the present debate, see, for example, among intentionalists, J. L. Bermudes, "SD, Intentions and Contradictory Beliefs," *Analysis,* 60, 2000, pp. 209–219.

[14] The causal view is best represented by Alfred Mele, *SD Unmasked,* Princeton University Press, Princeton, 2002.

[15] The issue of specificity is well addressed in the SD definition by D. Davidson, "Deception and Division," in E. LePore and B. McLaughlin, eds., *Actions and Events: Perspectives on the Philosophy of Donald Davidson,* Basil Blackwell, Oxford, 1985, pp. 138–148. Davidson's definition, however, gives rise to paradoxes, which need to be explained away by the problematic mind-partition. See the next chapter.

not well set apart from wishful thinking, on the one hand, or delusion, on the other. Similarly, there seems to be no problem for the intentional account in explaining why people confronting negative evidence do not always end up being self-deceived, given that SD is a doing of the subject.[16] If, by contrast, SD is caused by a motivational state taking a causal grip on mental tropisms, why does SD work *selectively* instead of being the normal response to wish-frustration? In sum, the causal account, for all its simplicity, does not provide convincing responses to the *specificity* and to the *selectivity* problems of SD. Therefore, in the critical reconstruction of the debate, I shall argue for a serious revision of the causal account enabling it to respond to these two major weaknesses, while subscribing to the nonparadoxical view of SD.

I shall then propose my solution of SD's puzzles in the form of a view of the phenomenon that is unintentional as to the outcome but acknowledges the intentionality of the process. In the realm of social explanations, the explanatory model accounting for unintentional consequences of intentional doings is the "invisible hand"; accordingly, I shall argue for an invisible hand account of SD, which will enable me to respond to the specificity issue while keeping a nonparadoxical view of the phenomenon. Similarly, the selectivity issue will be dealt with within the invisible hand model, but, in this case, the analysis shall be supplemented by a reflection on the circumstances conducive to SD. This will lead me to unpack the motivational state at the origin of the SD process. Finally, the analysis of the motivational state will help me sort out the issue of a unitary explanation of "straight" and "twisted" cases in the framework of the invisible hand model. The straight cases are those where there is a match between the wish and the false belief resulting from SD, as in the case of a husband who stubbornly refuses to believe his wife unfaithful despite evidence that would induce a nonmotivated observer to conclude the opposite. The twisted cases are those where the false belief runs counter to the agent's desire, as in the case of a jealous husband who becomes convinced of his wife's infidelity despite the lack of evidence. The intentionalists tend to deny that twisted cases are cases of SD. By contrast, the causalists claim to have provided a unitary explanation of straight and twisted SD, which would constitute a clear advantage over alternative explanations.[17] They hold that in both cases the self-deceptive belief is similarly produced by the

[16] See W. J. Talbott, "Intentional SD in a Single, Coherent Self," *Philosophy and Phenomenological Research*, 55, 1995, pp. 27–74, and Bermudes, "SD, Intentions and Contradictory Beliefs."
[17] See Mele, *SD Unmasked,* pp. 111ff.

causal working of a motivational state on cognitive biases, and whether the result is favorable or unfavorable to the subject's wish is irrelevant from the point of view of its production. In this perspective, the soothing effect of the false belief, typical of the more common straight cases, has nothing to do with the subject's desire to believe what she wishes but is simply a causal result, contingently going in one or another direction. This explanation, if correct, may undermine my account. I shall argue, however, that both cases are better explained within the invisible hand model as the outcomes of different strategies of reasoning, both affected by cognitive biases, chosen by subjects differently influenced by desires and emotions. A proper explanation of twisted cases as cases of SD is especially relevant in the political realm, for in this domain twisted cases often occur as the outcome of worst case scenario mode of reasoning. In situation of great uncertainty and risk, it is common to consider the worst as a precautionary move for working out potential responses. Such a move would not be intrinsically epistemically faulty, but it is usually exposed to the probability neglect bias, so that a quite improbable event, which represents the most feared outcome for the decision-maker, comes to be considered an actual probability. Given that the potential responses are necessarily quite drastic and harmful, the self-deceptive belief that worst case scenarios are real is very dangerous. Think, for example, of the reasoning backing the consideration of the potential nuclear capacity of Iran. In the uncertainty of reliable and accurate information, given the risk associated to it, many politicians and analysists thought it safer to consider it as a real possibility, and in fact believed that it was, instead of weighing the actual probability of such an event carefully. In the grip of such a belief, a preventive attack to Iran might look justified in view of its future nuclear potentiality. If, however, such a belief were a case of twisted SD, the justification of a preventive attack would fall apart, and the harmful and far-reaching consequences would be responsibility of the self-deceived decision makers.

In sum, the rather technical analysis on the three issues of specificity, selectivity and unitary explanation of both straight and twisted SD is relevant for the purpose of this book of applying SD to politics. A specific notion of SD is in fact required to detect its specimen precisely in the fog of political deception, by setting it apart from illusions, political myth, ideological assumptions and cold mistakes. The solution to the selectivity issue, furthermore, will help to define the circumstances where SD is likely to take place, and will provide the possibility to foresee SD and, hopefully, to devise preventive measure. Finally, the explanation of twisted cases, along with the more common straight cases, help